Understanding the relation between monitoring events and topology of exascale architectures for HPC applications

Idriss Daoudi, PhD

Argonne National Laboratory

December 15th, 2021

The Argo project

- **What?**
  - building low level system softwares for **resource management of exascale applications**
- **Why?**
  - improve the performance and scalability
  - provide new resource management **mechanisms** for exascale applications
- **How?**
  - provide new abstractions for resource management
  - configurable policies
  - dynamic application-aware resource management
  - portable, open source, validated, and scalability tested

Node Resource Manager (NRM)

- Daemon running on compute nodes
- Centralizes node management activities such as:
  - job management
  - resource management
  - **power management**

- Power management is **key for exascale era**:
  - allows to stay within the power budget
  - allows applications to make the most of the available power

- **Objective**:
  - balance complex applications requirements while keeping power consumption under budget

NRM: under the hood

- Application self-reporting: **progress**
  - processes use it to periodically update NRM on their progress
  - reliable feedback!

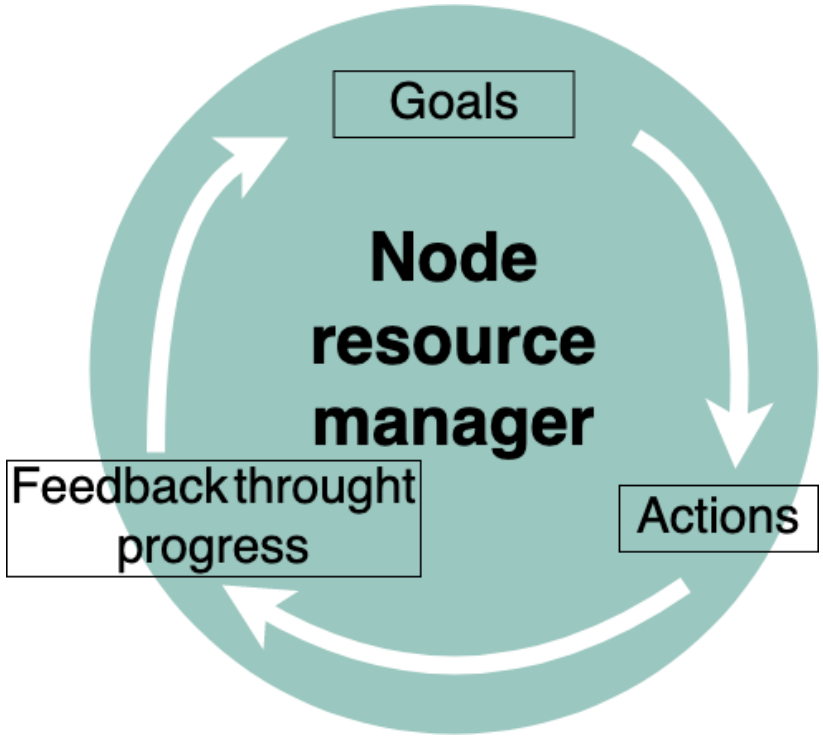
```
for (int i = 0; i < MAX; i++)
{
#pragma omp parallel for
    for (int j = 0; j < ITER; j++)
    {
        //some work
    }
    nrm_send_progress();
}
```

Figure: Example of an OpenMP application reporting progress

NRM: under the hood

- NRM works in a **closed control loop**:

1. set performance **goals**
2. **act** on applications workload
   - ★ adjust CPUs p-state
   - ★ modify powercap with Intel RAPL...
3. get feedback through **progress** and **monitoring**
   - ★ temperature
   - ★ frequency
   - ★ fan speed...



NRM: current advancements

- **Problem**
  - how to identify devices that are executing a certain process within an application?
- **Solution**
  - improvement of NRM sensor (monitoring) interface
- **Methodology**
  - identify monitoring events related to:
    - ★ the **location** (within the topology)
    - ★ the **scope** (range of devices)
  - apply improvements on hardware monitoring

What we are looking for

- We are aiming to **evaluate** our implementation
- **Observe** dynamic resource imbalance on complex applications
- **Address** it with a better power management strategy
- **Get a better understanding** of the behavior of such applications under **various scenarios** of power management
- Study the possibility of **characterizing** applications' power needs in order to develop an **automated** resource management policy

- Are you working on complex applications with dynamic resource balancing problems?
- Are you interested in such problematics?
- If yes, get in touch!

Acknowledgments

- This research was supported by the Exascale Computing Project (17-SC-20-SC), a joint project of the U.S. Department of Energy's Office of Science and National Nuclear Security Administration, responsible for delivering a capable exascale ecosystem, including software, applications, and hardware technology, to support the nation's exascale computing imperative.

- This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computer Research, under Contract DE-AC02-06CH11357.